Estimation of the Demand for Inter-city Travel: Issues with Using the American Travel Survey

Piyushimita Thakuria
Assistant Professor
College of Urban Planning and Public Affairs
Phone: (312) 355-0447 Fax: (312) 413-0006 Email: vonu-pt@uic.edu

Deepak Virmani
Research Assistant
Urban Transportation Center

Seongsoon Yun
Research Assistant Professor
Urban Transportation Center

Paul Metaxatos
Research Assistant Professor
Urban Transportation Center

University of Illinois at Chicago
412 South Peoria Street, Suite 340
Chicago, IL 60605

To be presented in

TRB Conference on

Personal Travel: The Long & Short of It

1. Introduction

Inter-city travel occurs for a variety of reasons. Decision makers need to have an estimate of the demand for long-distance travel in order to assess the level of service and the capacity of service for inter-city travel by different modes. Further, estimation of inter-city demand may serve to identify markets for new types of service as well as to deliver transportation services that meet the preferences of travelers. This problem is not unlike the intra-metropolitan area demand estimation that is routinely done by planning agencies in order to assess modifications/enhancements needed in the area's surface transportation infrastructure. But important differences exist.

The major objective of this paper is to demonstrate a methodological approach to estimating the pattern of long-distance highway travel demand (between large metropolitan areas), using data from the 1995 American Travel Survey (ATS)[1]. Our objective is not to develop a final model for estimating inter-city travel demand. It is rather to obtain an understanding of the types of costs that travelers consider in making long-distance destination choices as well as the nature of the statistical challenges that arise in estimating such demand using the ATS data. The approach used leads to several important bye-products. Foremost is the development of an approach that brings the estimation of inter-city travel demand by small area into the mainstream of travel demand modeling. Second, we would be able to estimate the changes in demand for travel between cities as a result of changes in costs of travel between cities. Further, the observed flow table (which gives the sampled counts of trips between each origin and destination metropolitan area) is likely to be `jagged' in the sense that a number of OD pairs may have zero counts, whereas others have large counts. The process allows the smoothing of the observed inter-city flow table, which is important for the ultimate purpose of prediction (prediction is not considered in this paper).

If we consider metropolitan areas across the United States as long-distance traffic producing and attracting zones, then flows between these origin-destination zone pairs can be estimated by gravity models of spatial interaction. This is a class of models that has been applied to estimate various types of transportation flows --- the most common application in transportation being to estimate origin-destination (O-D) flow within a metropolitan area. While the model has been in use for a long time, there has been a fair amount of recent developments in both the formulation of a behaviorally-based probabilistic model as well as in the numerical procedures that are used to estimate model parameters. For example, the theoretical basis of this type of model was recently examined in Sen and Smith (1995).

We have selected the fifty largest metropolitan areas for this problem. Also, we have considered only travel by private highway vehicle mode for the following reason. While the ATS dataset is extremely rich in terms of traditional demand-side variables (such as detailed household-level information on trips, trip purpose as well as the household socio-demographics), applications of travel demand models have large appetites for supply-side

---

[1] The ATS is a survey of long-distance travel and is conducted by the Bureau of Transportation Statistics, U.S. Department of Transportation.

variables, which are usually introduced into the models as costs. Except for distance traveled, other types of costs incurred by travelers in the ATS were not available from the dataset. Moreover, distances were available only for the observed inter-city trips in the ATS. But travel demand estimation requires distances between *all* (observed and unobserved) origin-destination pairs considered. Hence for the purposes of this paper, all cost or `impedance' data (distances and travel times) had to be synthetically generated, which proved to be a time-consuming, resource-demanding exercise.

Travel costs that travelers respond to for intra-urban trips are well understood at this time. This is not the case for the costs or impedances associated with inter-city travel. Hence the approach taken in this paper is exploratory. We first introduced into the model one impedance parameter and conducted statistical diagnostics, including examining good-of-fit measures. We then repeated the process for additional measures of impedance alternating between exploratory analysis, model specification, estimation and then diagnostics to understand the underlying patterns of inter-city trips. Further, travelers response to costs can be expected to vary by trip purpose, an issue we have considered in this paper.

Several unique challenges arise in using the ATS data to develop demand models. We have tried to address some of these challenges in this paper leaving the rest for future research. The paper is organized as follows: In Section 2, we briefly describe the data used. We describe the model proposed and the procedures used to estimate model parameters in Section 3. The special considerations that arise in using the ATS for demand models are described in Section 4. We present different scenarios for inter-city travel demand and the scenario estimation results in Section 5. Finally, we present our conclusions in Section 6.

2. Data Used for the Inter-City Problem

ATS is a comprehensive source of data on the inter-city travel characteristics of the people in USA. This survey was carried out in 1995 on a sample of 80,000 addresses selected from the 1980 base Current Population Survey (CPS) sample. The sample was selected to ensure proportional representation of travel patterns of people in all the states. Apart from the trip information the database also includes the demographic and socio-economic characteristics of the surveyed sample. ATS also has an extensive system of weights which when applied provide an estimate of all the inter-city travel that took place in the country in one whole year for different purposes and by different modes.

ATS data can be analyzed at an aggregated level for the whole country or by census regions and divisions or for individual states. At a more disaggregated level it can be analyzed for the metropolitan areas – both Metropolitan Statistical Areas (MSAs) and Primary Metropolitan Statistical Areas (PMSAs). The metropolitan areas (MSAs and PMSAs) designated in the ATS are not necessarily the same as those currently defined by the Office of Management and Budget (OMB) but are those with estimated 1995 population of 250,000 or more.

2.1 ATS Data Used

The ATS data are available in the form of a household and a person trip files. Both these files contain similar trip information. For our present research we have used the person trip data which contains some additional information (income, activity, age etc.) about the individual trip maker.

The person trip file contains almost 348 different variables but we selected only a small subset of the trip file as an input in to the gravity model for estimating the inter-city travel. The trips selected for modeling were personal use vehicle trips at the Metropolitan Area level. The personal use vehicle trips as defined by the ATS are those in which the main type of transportation used to cover most of the miles on the trip was auto, pickup truck, van, other truck, rental car, truck or van, recreational vehicle or motor cycle. We selected only those trips where these modes were used for the round trip.

Although there are 162 different metropolitan areas contributing to the ATS trip sample, we considered the trips between 50 of these origin-destination pairs. These areas are presented in Table 1. The criteria for selection were the area's population size and the area's occurrence as a trip origin node in the sample. Additionally, we ensured that all the census regions in the country were well represented. The population estimates used for this purpose were the 1996 Bureau of Census estimates for the metropolitan areas. The selected metropolitan O-D pairs are mentioned below in an alphabetical order.

1	Albuquerque, NM MSA	26	Milwaukee-Waukesha, WI PMSA
2	Atlanta, GA MSA	27	Minneapolis-St. Paul, MN MSA
3	Austin-San Marcos, TX MSA	28	Nashville, TN MSA
4	Baltimore, MD PMSA	29	New Orleans, LA MSA
5	Boston, MA PMSA	30	New York, NY PMSA
6	Charlotte-Gastonia, NC MSA	31	Norfolk-Virginia Beach-Newport News, VA MSA
7	Chicago, IL PMSA	32	Oakland, CA PMSA
8	Cincinnati OH-KY PMSA	33	Oklahoma City, OK MSA
9	Cleveland-Lorain-Elyria, OH PMSA	34	Omaha, NE MSA
10	Colorado Springs, CO MSA	35	Philadelphia, PA-NJ PMSA
11	Columbus, OH MSA	36	Phoenix-Mesa, AZ MSA
12	Dallas, TX PMSA	37	Pittsburgh, PA MSA
13	Denver, CO PMSA	38	Portland-Vancouver, OR-WA PMSA
14	Detroit, MI PMSA	39	Sacramento, CA PMSA
15	El Paso, TX MSA	40	Salt Lake City-Ogden, UT MSA
16	Fort Worth-Arlington, TX PMSA	41	San Antonio, TX MSA
17	Fresno, CA MSA	42	San Diego, CA MSA
18	Houston, TX PMSA	43	San Francisco, CA PMSA
19	Indianapolis, IN MSA	44	San Jose, CA PMSA
20	Jacksonville, FL MSA	45	Seattle-Bellevue-Everett, WA PMSA
21	Kansas City, MO-KS MSA	46	St. Louis, MO-IL MSA
22	Las Vegas, NV MSA	47	Tucson, AZ MSA
23	Los Angeles-Long Beach, CA PMSA	48	Tulsa, OK MSA
24	Memphis, TN MSA	49	Washington, DC-MD-VA PMSA

25	Miami, FL PMSA	50	Wichita, KS MSA

Table 1: Metropolitan areas considered in the inter-city flow estimation problem.

2.2 Creation of Impedance Data

Cost parameters used in intra-urban travel demand modeling typically include distance between origins and destinations and travel time. The ATS data set gives the highway route distance between origin and destinations. However, the data for this variable in the dataset could not be used for the gravity model application because the model requires the distance between *all origin-destination pairs* considered in the study as input. If none of the respondents in a particular metropolitan area undertook a trip to a particular destination, then there would be no distance data available for that particular O-D pair. Hence, we had to look at alternative sources for data on distances between metropolitan areas. There were no data on travel times between O-D pairs in the survey at all.

Creating the cost matrices for inter-city travel demand estimation turned out to be a non-trivial and time-expensive effort. We gathered distances and travel times provided by several web-based databases and selected one on the basis that the travel times and distances reported appeared to be reasonable. The database used has been created by ETAK, Inc. (part of Sony Corp.) a California based company which specializes in navigation technologies and geocoding. The data we procured from their vendor's web site provides information on the distance along the shortest path between an O-D pair and travel times on those paths. The travel times are based on posted speed limits on different roadway facilities along the route and does not include congestion effects in any way.

3. Gravity Model of Inter-city Flows

We estimated patterns of demand between the fifty different metropolitan areas using gravity models. In this section, we describe the model estimated.

Let $I = \{1,2,\ldots\ldots I\}$ be a set of origin metropolitan areas, $J = \{1,2,\ldots\ldots,J\}$ be a set of destination metropolitan areas and $c_{ij} = (c_{ij}^{(1)}, c_{ij}^{(2)}, \ldots, c_{ij}^{(K)})$ be a set of K costs, such as travel time, distance and so on, which separates $i \in I$ from $j \in J$. We consider a gravity model of inter-city flows which is given by

$$E(N_{ij}) = T_{ij} = A_i B_j F(c_{ij}) \qquad : \qquad i \in I,\ j \in J \qquad (1)$$

where T_{ij} is expected flow between zone i and j. A_i's and B_j's are origin and destination functions respectively and

$$F(c_{ij}) = \exp[\theta^t c_{ij}]. \qquad (2)$$

The vector $\theta^t = (\theta_1, \theta_2,\ldots\ldots,\theta_K)$ includes the parameters associated with the separation measures c_{ij}. While various functional forms for $F(c_{ij})$ are possible, the exponential form is general enough for most applications.

The gravity model gives estimates of expected aggregate flows between origins and destinations and thus is ideally suited for inter-city travel demand estimation. The end result (after estimation) is an estimated "trip table" containing t_{ij}'s (t_{ij} is the estimate of T_{ij}), the estimated zone-to-zone flow, which is an assessment of inter-city demand.

Once the estimates of θ are available, it would be possible to assess the changes that would occur in the resultant flows, if costs of travel between metropolitan-area pairs were to change. Further, as discussed in Section 1, we could use the estimate of θ to estimate flows between smaller areas. For this latter purpose, all we would need are trip production and attraction rates by small areas. Using these trip generation numbers and the estimated value of θ, we could construct trip tables of estimated flows between one smaller zone to another. If done properly, the ATS data could then yield estimated flows by smaller geography.

3.1 Inter-city Model Parameters

Earlier we referred to a vector of parameters (θ), of impedance to travel between origins and destination cities or regions. Our approach in this paper has been to fit a model using one estimated impedance parameter first, examine the residuals and then estimating another impedance parameter to explain the residuals. This is an "iterative" process, but it allows us to avoid overfitting the model with too many costs parameters and consequently, to avoid the costly process of developing additional cost data. At the same time, the process of estimation, model fitting and examination of the patterns in the residuals also allows us to obtain insights on the types of separation measures that travelers respond to, in determining a destination for inter-city travel. These separation or impedance measures relevant to inter-city travel may not be apparent *a priori*.

In Equation (1), A_i, B_j and θ_K are the parameters to be estimated from available data. Once the θ_K's are estimated [by Maximum Likelihood (ML)], we obtain estimates of A_i's and B_j's using a variant of Iterative Proportional Fitting. The A_i's and the B_j's are non-unique (they are unique conditional on the θ's). The estimation procedures are given in Section 3.2.

3.2 Estimation Procedures

The procedures to estimate the θ_k's have been studied extensively (see Sen and Smith, 1995). However, for the sake of completeness, we briefly review the procedures used in this paper.

Each random variable, N_{ij}, of flows between i and j, can be assumed to be independently Poisson-distributed such that

$$P(N_{ij}) = e^{-T_{ij}} T_{ij}^{N_{ij}} / N_{ij}\,! \tag{3}$$

Thus the probability of observing the trip pattern **N** is:

$$P(\mathbf{N}) = \prod_{ij} P(N_{ij}) = \prod_{ij} \exp[-T_{ij}]\, T_{ij} \;/\; N_{ij}\,!$$

$$= \prod_{ij} \{\exp(-A(i)B(j)\exp[\theta^{T} c_{ij}])(A(i)B(j)\exp[\theta^{T} c_{ij}])^{N_{ij}} / N_{ij}\,!\}. \qquad (4)$$

Since the N'_{ij}s are known from the ATS data and Equation (4) can be viewed as a function of the parameters A_i's, B_j's and the θ_k's. Equation (4) in fact is a likelihood function. The values of the estimates of A_i's, B_j's and the θ_k's that maximize this function (or the logarithm of this function) are the Maximum Likelihood estimates. Taking the partial derivatives of the logarithm of Equation (4) with respect to A_i, to B_j and to θ_k and setting each case equal to zero, we get the sequence of equations:

$$\sum_{j=1}^{J} T_{ij} = \sum_{j=1}^{J} N_{ij} \;\text{ for all } i \in I, \qquad \sum_{i=1}^{I} T_{ij} = \sum_{i=1}^{I} N_{ij} \;\text{ for all } j \in J \qquad (5)$$

$$\text{and } \sum_{ij} c_{ij}^{(k)} T_{ij} = \sum_{ij} c_{ij}^{(k)} N_{ij} \;\text{ for all } k \in K. \qquad (6)$$

Numerous procedures have been developed recently to allow Equations (5) and (6) to be solved in a computationally efficient way (Yun and Sen, 1994). These equations are obviously nonlinear. We have used the Modified Scoring Procedure which includes procedures to linear approximations at each iteration, using a method called the Linearized Deming-Stephan-Furness (LDSF) procedure, which is a variant of the well-known Iterative Proportional Fitting approach. The interested reader is referred to Sen and Smith (1995) for detailed discussions on this topic.

4. Unique Characteristics of the ATS: Implications for Demand Estimation

Before we go to the models estimated in this paper, we note our findings regarding the special challenges that arise with the use of the ATS data for inter-city demand estimation. Some of these were alluded to in Sections 1 and 2.

A. **Need to develop cost matrices from exogenous sources:** Various procedures in the traditional demand-modeling framework, including the gravity model, work with the concept of “costs” to travel. Traditionally, for origin-destination flow estimation purposes, impedances such as distance and travel time are used. These are not directly available from the ATS dataset (as discussed above in Section 2.2). Hence, important cost data creation and linking issues arise before the ATS demand-side data can be used for demand estimation purposes.

B. **Time scale:** Typically, in intra-urban modeling exercises, travel demand is estimated over a day (or over different time periods of a day). This is the better temporal unit to consider in intra-urban situations, in assessing the need for additional capacity, especially for highway networks and is especially important for the assignment stage of a demand modeling exercise. The ATS ultimately gives annual volumes between origin-destination pairs. To be useful in assessing capacity needs, there is the issue of "scaling" the ATS demand data to a finer time resolution.

C. **Spatial aggregation:** There is also a need to avail of inter-city data at a finer spatial resolution (by small areas). For example, the ATS currently does not give the analyst the ability to distinguish one airport from another, if there are two airports in the same metropolitan area. However, in order to assess issues of ground access control in major airports and train terminals, a situation which would be of increasing importance with increasing demand for inter-city travel, data are necessary by smaller areas. Further, infrastructure planning is handled by different jurisdictions within a metropolitan area and hence estimates of travel demand, by each smaller jurisdictions, become important to obtain.

5. Scenarios Evaluated and Results

As discussed earlier, we approached the problem of estimating inter-city flows with the intention of understanding which parameters allow the best replication of observed flows. In this section, we present the various scenarios that were modeled using the ATS data. The scenarios involve both single and multiple variable parameter estimation cases. In addition, one set of scenarios consider trip purpose as conditioning information. Table 2 gives the names of the scenarios and the type of trips considered as well as the cost parameters included in each case.

Scenario	**Type of trips**	**Trip Type**	**Cost Parameters**
Case IA	All trips	AT*	distance
Case IB	All trips	AT	travel times
Case IC	All trips	AT	distance and travel times
Case IIA	visit relatives or friends rest or relaxation sightseeing outdoor recreation entertainment, shopping personal, medical or family	NBT**	distance
Case IIB	sightseeing & outdoor recreation	RT***	distance
Case IIC	sightseeing & outdoor recreation	RT	distance and travel times
Case IID	sightseeing & outdoor recreation	RT	distance and log of distance

* AT: All Trips ** NBT: Non Business Related Trips *** RT: Recreational Trips

Table 2: Scenario names, type of trips and cost parameters of inter-city demand models.

The discussion on the results highlight the various challenges associated with estimating inter-city demand from the ATS data. The challenges include the need for better cost data as well statistical problems that need to be addressed. For example, it may be necessary to construct meaningful indices to model attraction of flows between originating and destinating metropolitan areas. For business-related trips for instance, it may be necessary to construct an index that gives the proportions of an area that is in different occupational categories. For non-business related trips, there may be a need to explore the creation of indices that take into account sales and other tax structures to explain shopping trip flows, and migration patterns to explain trips with the purpose of visiting friends and relatives. These challenges was alluded to in Section 4. We gain some insight into the need for more detailed cost data in Section 5.2.

Inspite of adequate cost data, certain statistical problems may remain. Multicollinearity in multiple variable models may be a problem. We see the possibility of this problem in Cases IC, IIC and IID. Further, the models may lead to estimates with high variance. A way to reduce the variance of estimates is to estimate conditional expectations. However, in the process of adding a conditioning structure in the model, we lose sample size, resulting, in some cases, in extremely sparse observed trip matrices. Although cases with extremely sparse matrices is theoretically not a problem with the gravity model, there is a potential of facing instability in the numerical procedures. Instances of these issues are highlighted in Sections 5.1 and 5.2.

5.1 Scenario I: All Trips

Under this scenario, we considered all trips [AT] between the 50 city pair, irrespective of trip purpose. In Case IA, the only impedance variable considered is distance. In Case IB, travel times between metropolitan areas were considered. In Case IC, both distance and travel times were considered. The models in all three cases thus contain 50 unknown origin parameters and 50 destination parameters. The model in Cases IA and IB includes one parameter for the separation measures considered in each case, whereas the model in Case IC includes two separation measure parameters. The parameters, including θ_k for separation measure $c^{(k)}$ for each origin-destination are estimated using Maximum Likelihood estimation, which was discussed in Section 3.2. Also, as discussed in Section 3.1, the origin and destination parameters are not unique. However, fixing one of them arbitrarily renders the remaining parameters unique. The estimates of T_{ij} are unique (Sen and Smith, 1995).

Parameter Scenario	Estimate	χ^2 *	χ^2 /df **
IA [AT]			
distance	-0.00187	1107236.00	461.35
IB [AT]			
travel time	-0.00196	1056551.00	440.23
IC [AT]			
distance	-0.00134		
travel time	-0.00336	1059256.00	441.54

* $\chi^2 = \Sigma [N_{ij} - t_{ij}]^2 / t_{ij}$

** df = (I-1)(J-1)(K-1) where I = Number of Origins, J = Number of Destinations, and K = Number of Parameters

Table 3. Scenario I results: Parameter estimates and goodness-of-fit measures for all private vehicle trips.

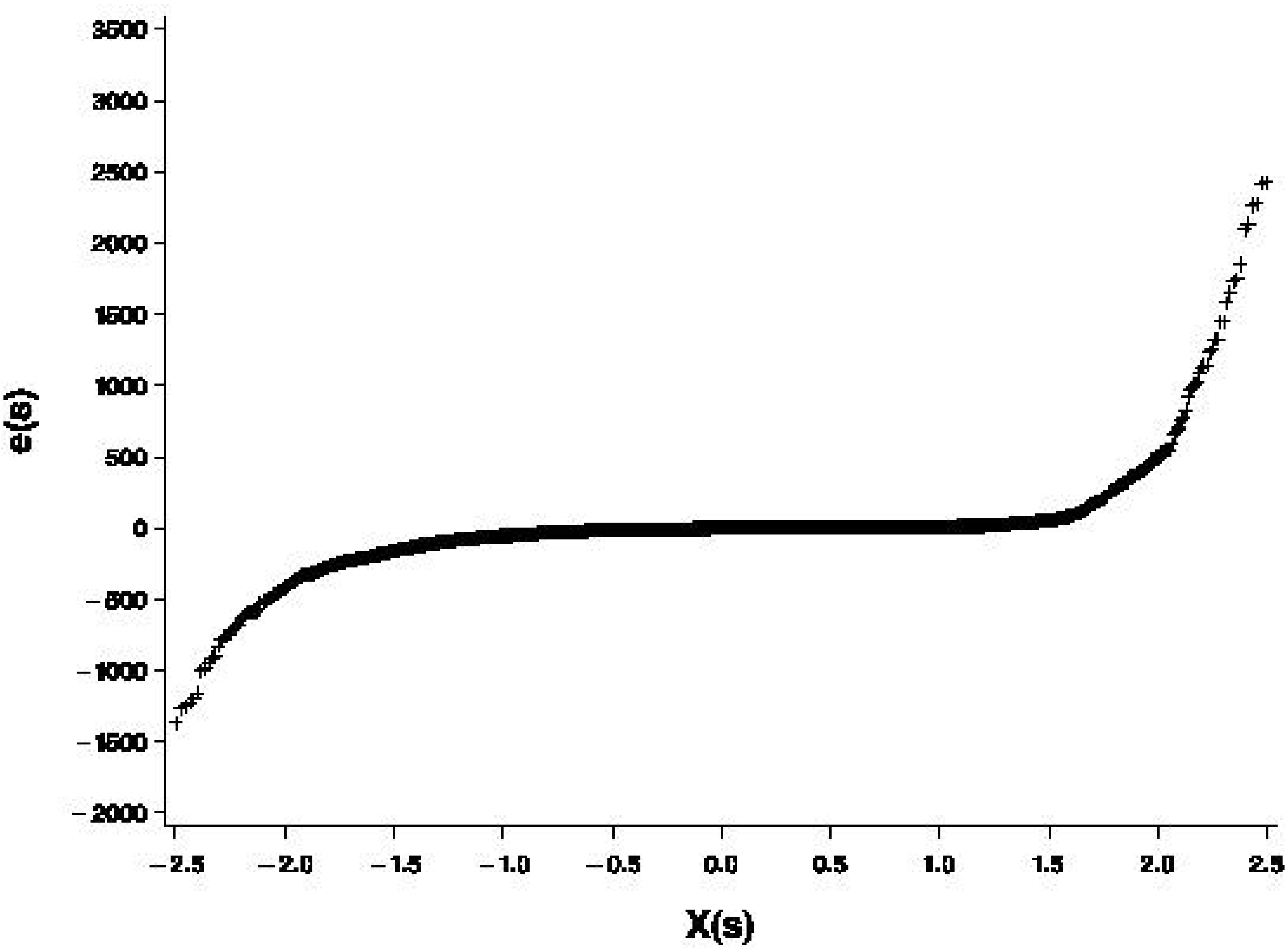
Table 3 gives the cost parameter estimates, the chi-squares (χ^2) and the chi-square ratios (χ^2 / df) for the three different cases. Goodness of fit of gravity models are often evaluated using the chi-square statistic which, with a large number of observations, should equal the degrees of freedom if the individual trips between cities are independent. Since there is not complete trip independence (in the sense that flows between an origin and a destination may be composed of joint trips by members of a household and so on), we would expect a chi-square ratio (the chi-square statistic divided by the degrees of freedom or χ^2/df, which is given in the fourth column of Table 3) of greater than 1. Models with chi-square ratio values closer to 1 indicate better fits. Although the chi-square ratio of Case IB, with travel time only, is lower than that in Case IA, with distance only, the difference is not too great. Also, using both travel time and distance together as in Case IC, leads to an increase in the chi-square ratio, although the increase in negligible.

One problem that may arise with using distance and travel time together in the inter-city case is the possible presence of multicollinearity. Although the presence of multicollinearity could seriously affect the variances of the estimates, typically, in the intracity case, deletion of travel times may bias the estimates, which would occur by leaving out an important variable from the model. In the intracity case, distances and travel times are not necesarily linearly related because of the presence of congestion effects. In the inter-city case, however, the only congestion that drivers on long distance trips incur would probably be in urban areas enroute. There is a possibility that drivers make up the time lost in congested urban areas by driving faster in the rest of the trip thus bringing about a “smoothing” effect in total trip times. This would tend to occur especially on long trips.

But inspite of the possible presence of the “smoothing” effect in total trip times, care still needs to be taken to evaluate the relationship between distances and travel times. Although we have reason to suspect multicollinearity in Case IC, notice that we had used travel times that does not include congestion effects (see Section 2.2). Congestion, which would be present at least in large metropolitan areas, could cause travel times and distances not to be linearly related at least for origin-destination pairs within short distances metropolitan corridors. Hence it may be necessary to consider (i) travel times that reflect congestion in urban areas and (ii) the inclusion of additional terms such as indicator variables based on distance, that allows the analyst to include the effect of the smoothing effect into the model. But at this time, we do not have these congested travel times. Further, it intuitively seemed to us that distance would be the foremost variable on the basis of which travelers undertake inter-city highway trips. Hence, we decided to exlude travel time from the rest of the analysis. This would take care of possible multicollinear effects and still allow us to avoid bias in the estimates.

Figure 1

Case IA: All trips with distance cost parameter



Rankit plots of the residuals of the model in Case IA [that is, the sorted residuals of the quantity [N_{ij}-t_{ij}], e(s) where (s) is the sorted order of the residual, against X(s) = φ^{-1}[(s-3/8)(n+1/4)], where φ is the distribution function (cdf) of the standard normal distribution and n=IJ is the total number of residuals] pointed to the presence of possible outliers in the estimated model. The plot is shown in Figure 1. These were flagged and after further scrutiny, found to be largely for cases of origin-destination pairs which are less than 500 miles or so. The fit for origin-destinations greater than 500 miles appeared to be very good. The presence of `jaggedness' in the plots combined with a nonlinear appearance suggested that appropriate transformations of the cost variable be included in the model. We considered several of these and the results of promising transformations are given for the models considered under Scenario II. Further, the size of the residuals suggested that some sort of conditioning structure be used to scale down the variance.

5.2 Scenario II: Model by Trip Purpose

It is reasonable to expect that the pattern of inter-city trips would vary by the purpose of trips. Under this scenario, we consider what we have called non-work trips. Case IIA includes a broad category of Non-Business related Trips [NBT]. These included the categories of (1) visit relatives or friends (2) rest or relaxation (3) sightseeing (4) outdoor recreation (5) entertainment (6) shopping (7) personal, medical or family in the ATS data. Therefore, under this scenario, we are estimating a conditional expectation, the conditioning information being trip purpose.

Figure 2(A)

Case IIA: Non – Business Trips with distance cost parameter

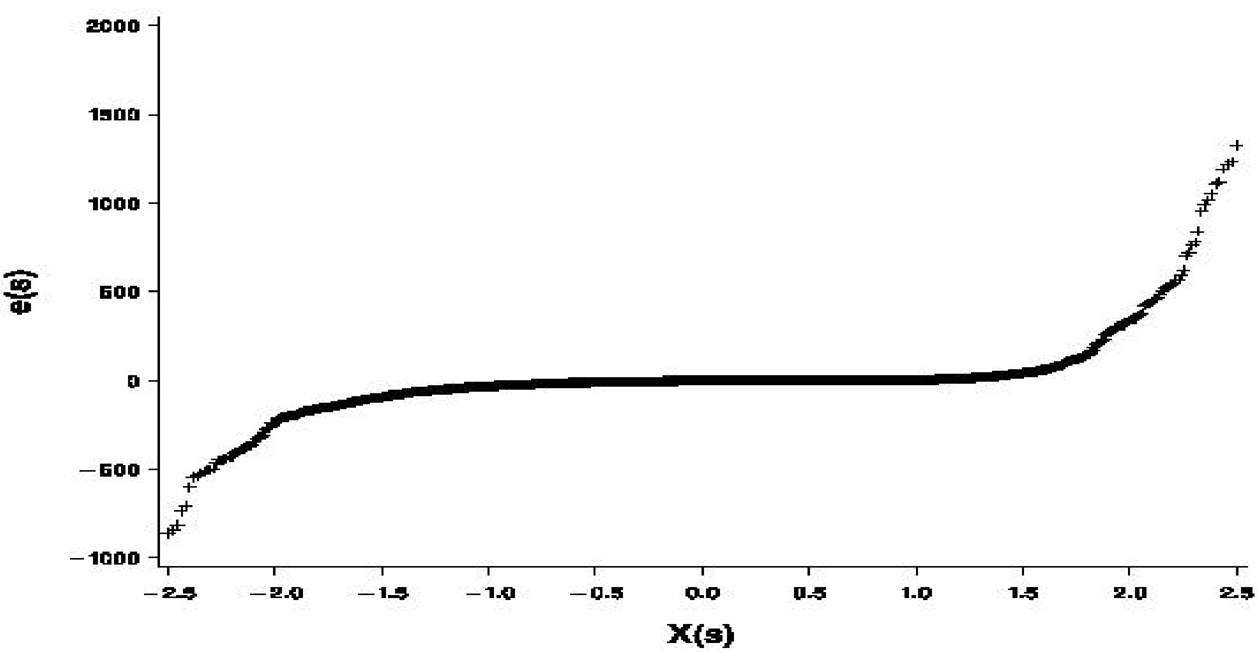
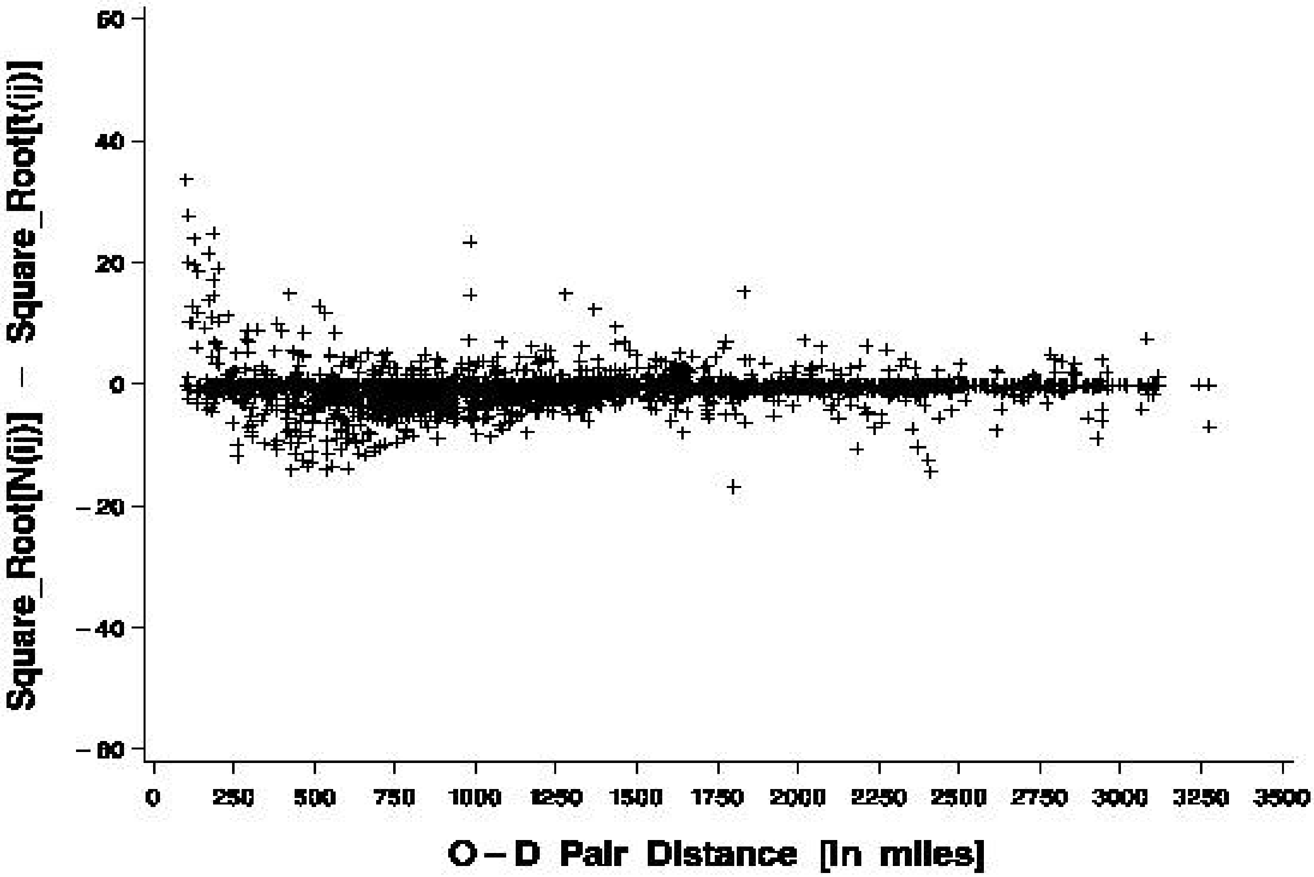


Table 4 gives the estimate of the distance parameter and the associated chi-square statistic and chi-square ratio. It can be seen that the fit has improved in case IIA [with χ^2/df=388.50 compared to χ^2/df=461.35 for Case IA, the corresponding case for all trips]. However, rankit plots of case IIA [given in Figure 2(A)] still suggested the presence of large residuals.

Figure 2(B)

Case IIA: Non – Business Trips with distance cost parameter



Visual inspection of plots of residuals [N_{ij}-t_{ij}] against distance showed that the large residuals are mostly for trips below 500 miles. But these O-D pairs also higher flows. In order to deal with this artifact of the Poisson distribution, we considered a transformation of the residuals, namely the difference between the square root of N_{ij} and the square root of t_{ij} . This quantity is plotted against distance in Figure 2(B) and it shows that there is virtually no pattern left in the residuals except for very short trips of less than 150 miles or so, indicating that the model fits the observed data adequately.

Scenario	Parameter Estimate	χ^2	χ^2 /df
IIA [NBT]			
Distance	-0.00188	932378.94	388.50
IIB [RT]			
Distance	-0.00189	560058.69	233.36
IIC [RT]			
Distance	-0.00189		
Travel Time	-0.00189	522982.69	218.00
IID [RT]			
Distance	-0.00523		
Log of Distance	-0.23293	604548.00	252.00

Table 4: Scenario II results for all non-work trips: Parameter estimates and goodness-of-fit measures.

Case IIB includes only Recreational Trips [RT] under which we have included the sightseeing and outdoor recreation categories of trips in the ATS data. The category Non-Business related Trip purposes [NBT] considered under Case IIA would appear to intuitively lead to different travel patterns than trips made simply for the purpose of sightseeing and recreation. For example, shopping trips to a destination may be related to sales or other tax situations whereas visiting friends and relatives may follow migration patterns that would require us to include more complex matching indices between origins and destinations.

Further, Cases IIB, IIC and IID are presented here for the purpose of highlighting the types of statistical and numerical problems that arise with imposing a strict conditioning structure on the inter-city flow estimation problem. In the process of conditioning by further breakdown of trip purposes, we appeared to have compromised on, in Cases IIB, IIC and IID, sample size. The cell counts for the observed trip matrix turned out to be very small. In the RT case, more than 96% of the trip table had zero-valued cells. This caused various numerical instabilities in the estimation procedures. This means that the estimates of θ_k did not necessarily converge to their true values, leading to instabilities in the numerical outputs, including the goodness-of-fit measures. Further, at least with Cases IIC and IID mutlicollinearity may remain a problem.

In Case IIB, recreational trips were estimated with a single cost parameter of distance. While the chi-square ratio improves substantially [to χ^2/df =233.36] in Case IIB, the rankit plot of this case suggested that additional variables may be necessary. In Case IIC therefore, we considered, in addition to distance, travel times. As discussed under Case IC, since multicollinearity would affect variances of estimates, the use of distance and travel time together for the inter-city case would not be advisable for applications relating to predictions, especially for origins and destinations that are located at great distances from each other. However, it is useful to include both variables to see how much the fit improves. The fit improved with the inclusion of travel time [to χ^2/df =218] but it is only marginally better than Case IID [χ^2/df=252], where we included distance and log of distance. In fact, the three cases, IIB, IIC and IID, of recreational trips, result in similar fits. However, the numerical instability issue with the RT cases need to be kept in mind and serves to remind the analyst that various trade-offs may need to be done between avoiding linear dependency in the cost matrices, higher variance and instability of estimation procedures.

6. Conclusions

In this paper, we used the American Travel Survey to demonstrate a methodological approach to estimating the pattern of long-distance highway travel demand (between large metropolitan areas), using data from the 1995 American Travel Survey (ATS). Our objective was to obtain an understanding of the types of costs that travelers consider in making long-distance destination choices as well as the nature of the statistical challenges that arise in estimating such demand using the ATS data. We considered travel data between fifty (50) metropolitan areas.

The paper used a gravity model of spatial interaction to estimate demand for inter-city metropolitan area. The cost parameters were estimated by Maximum Likelihood and the origin and destination parameters were estimated using a variant of Iterative Proportional Fitting, which we call the Deming-Stephan-Furness (DSF) procedures. Various scenarios were estimated using the ATS data, including a model of all highway trips between the 50 metropolitan area pairs as well as models for specific trip purposes.

An important outcome of this model is the development of an approach that allows the estimation of inter-city travel demand by small area into the mainstream of travel demand modeling. Further, the observed flow table (which gives the sampled counts of trips between each origin and destination metropolitan area) is likely to be `jagged' because a number of OD pairs may have zero counts, whereas others have large counts. The process allows the smoothing of the observed inter-city flow table, which is important for the ultimate purpose of prediction.

The modeling exercise served to highlight various challenges associated with estimating inter-city demand from the ATS data. The challenges include the need for better cost data as well statistical problems that need to be addressed.

Several important cost data creation and linking issues arise before the ATS demand-side data can be used for demand estimation purposes. It may be necessary to construct meaningful indices to model attraction of flows between metropolitan areas. The cost data needed for estimating inter-city demand are not directly available from the ATS dataset and needs to be contructed from exogenous sources. Further, in order to assess “capacity” issues between city pairs, annual volumes between origin-destination pairs, there is the issue of “scaling” the ATS demand data to a finer time resolution. There is also a need to avail of inter-city data at a finer spatial resolution (by small areas) in order to assess issues of ground access conditions in specific inter-city trip originating and ending points.

Even if the cost data were adequate, we may still encounter complicated statistical problems in estimating inter-city demand one of which is multicollinearity in multiple variable models. Further, the models may lead to estimates with high variance. However, by imposing a conditioning structure to reduce variance we lost sample size, resulting, in some cases, in extremely sparse observed trip matrices. Although cases with extremely sparse matrices is theoretically not a problem with the gravity model, we encountered instability in the numerical procedures used to estimate some of the parameters. These findings served to remind us that various decisions may need to be made to avoid multicollinearity among the cost variables in the model, to lower the variance of estimates and to obviate instability in the numerical procedures.

References

1. Bureau of Transportation Statistics (1995). American Travel Survey. U.S. Department of Transportation, Bureau of Transportation Statistics, Washington, D.C.

2. Etak, Inc. (1998). *EtakMap Premium* . Menlo Park, California.

3. Sen, A. and T.E. Smith (1995). *Gravity Models of Spatial Interaction Behavior*. Springer-Verlag, New York.

4. Yun, S. and A. Sen (1994). Computation of Maximum Likelihood Estimates of Gravity Model Parameters. *Journal of Regional Science* 34, 199-216.